REVIEW

What do we know about grant peer review in the health sciences? [version 1; referees: 1 approved, 1 approved with reservations]

Susan Guthrie[1], Ioana Ghiga[1], Steven Wooding[2]

[1]RAND Europe, Westbrook Centre, Milton Road, Cambridge, UK
[2]Centre for Science and Policy, University of Cambridge, Cambridge, UK

v1 **First published:** 07 Aug 2017, **6**:1335 (doi: 10.12688/f1000research.11917.1)
Latest published: 07 Aug 2017, **6**:1335 (doi: 10.12688/f1000research.11917.1)

Abstract

Background: Peer review decisions award >95% of academic medical research funding, so it is crucial to understand how well they work and if they could be improved.

Methods: This paper summarises evidence from 105 relevant papers identified through a literature search on the effectiveness and burden of peer review for grant funding.

Results: There is a remarkable paucity of evidence about the overall efficiency of peer review for funding allocation, given its centrality to the modern system of science. From the available evidence, we can identify some conclusions around the effectiveness and burden of peer review.

The strongest evidence around effectiveness indicates a bias against innovative research. There is also fairly clear evidence that peer review is, at best, a weak predictor of future research performance, and that ratings vary considerably between reviewers. There is some evidence of age bias and cronyism.

Good evidence shows that the burden of peer review is high and that around 75% of it falls on applicants. By contrast, many of the efforts to reduce burden are focused on funders and reviewers/panel members.

Conclusions: We suggest funders should acknowledge, assess and analyse the uncertainty around peer review, even using reviewers' uncertainty as an input to funding decisions. Funders could consider a lottery element in some parts of their funding allocation process, to reduce both burden and bias, and allow better evaluation of decision processes. Alternatively, the distribution of scores from different reviewers could be better utilised as a possible way to identify novel, innovative research. Above all, there is a need for open, transparent experimentation and evaluation of different ways to fund research. This also requires more openness across the wider scientific community to support such investigations, acknowledging the lack of evidence about the primacy of the current system and the impossibility of achieving perfection.

Open Peer Review

Referee Status: ? ✓

	Invited Referees 1	2
version 1 published 07 Aug 2017	? report	✓ report

1 **Adrian Barnett**, Queensland University of Technology, Australia

2 **Aled Edwards**, University of Toronto, Canada

Discuss this article

Comments (0)

Corresponding author: Susan Guthrie (sguthrie@rand.org)

Author roles: Guthrie S: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ghiga I**: Data Curation, Formal Analysis, Investigation, Project Administration, Writing – Original Draft Preparation; **Wooding S**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Guthrie S, Ghiga I and Wooding S. **What do we know about grant peer review in the health sciences? [version 1; referees: 1 approved, 1 approved with reservations]** *F1000Research* 2017, **6**:1335 (doi: 10.12688/f1000research.11917.1)

Grant information: This report was produced with funding from the Canadian Institutes of Health Research.

First published: 07 Aug 2017, **6**:1335 (doi: 10.12688/f1000research.11917.1)

Introduction

Health research has contributed enormously to society, but it is also expensive. This has led to increasing demands to understand and improve how research is supported. Most effort has focused on evaluating impacts of research, on society and the economy. Funders are gathering evidence of impact using online survey platforms such as Researchfish in the UK, and national assessment frameworks including Excellence for Research in Australia (ERA).

Much less work has focussed on understanding how research is selected for support. Peer review is used to allocate the vast majority of competitive research funding internationally. Therefore it is crucial to understand whether peer review is effective and efficient - whether it can fairly, reliably allocate research funding without bias. In this study, we carried out a rapid evidence assessment which asked whether the peer review process lives up to these aspirations.

The research was commissioned by the Canadian Institutes of Health Research (CIHR) to support an ongoing review of CIHR's peer review system, particularly the Peer Review Expert Panel which was convened to review the design and adjudication processes of CIHR's investigator-initiated research programmes.

Methods

Search strategy

We identified relevant literature through five routes:

1. Google Scholar search using the search terms below, for publications from 2009 onwards. We reviewed the top 500 search results for each query.

Search terms:

- 'Grant peer review'
- 'Grant review' AND 'panel'
- ('Peer review' AND 'funding application') OR ('peer review' AND proposal) OR ('peer review' AND funding) OR ('peer review' AND award) or ('peer review' AND 'reviewer bias')

2. Grey literature: we searched the websites of major funding bodies and other academic bodies (e.g. learned societies) that we expected to have published relevant research (Table 1).

3. Searching the Cochrane publication list for systematic reviews on grant peer review. This did not identify any relevant reviews conducted since 2009.

4. An initial set of publications already known to the authors and sponsors of the work.

5. Snowballing: from the reference lists of publications identified following screening.

Table 1. Academic bodies considered in the review of literature.

Organisation	Country
National Institutes of Health (NIH)	USA
Canadian Institutes of Health Research	Canada
Health Research Board of Ireland	Ireland
Science Foundation of Ireland	Ireland
Netherlands Organisation of Health Research and Development (ZonMw)	Netherlands
Research Council of Norway	Norway
National Institute for Health Research (NIHR)	UK
Wellcome Trust	UK
National Health And Medical Research Council	Australia
Health Research Council of New Zealand	New Zealand
Medical Research Council	UK
Deutsche Forschungsgemeinschaft (DFG)	Germany
Lundbeck Foundation, Copenhagen	Denmark
Swedish Medical Research Council	Sweden
Swedish Society for Medicine	Sweden
European Commission	European Union

Some elements of our strategy were focused on evidence from the health sciences (particularly grey literature), but our wider searches, including Google Scholar, were not restricted by field of research.

Screening strategy

Publications were initially screened on title, and abstract (where available). Studies needed to include empirical consideration of the effectiveness and/or burden of grant review processes. Studies were excluded on the basis of being:

- Purely descriptive, describing a specific peer review process.
- Focused on wider concerns around the funding process, with no (or only tangential) reference to the peer review process in particular.
- Focused on manuscript peer review rather than peer review for funding purposes.
- From 2008 or earlier.
- Reviews, with no additional synthesis or analysis, summarising work from before 2008, or studies already identified and included individually.

If studies were relevant full text was retrieved and an Excel spreadsheet was used to capture key information on the study and its conclusions.

We identified 105 studies for inclusion. Table 2 summarises the range of studies identified.

Assessment of evidence quality

Quality of evidence was rated on a scale of 1–4 based on GRADE (Guyatt *et al.*, 2008)[1]. We aggregated the overall strength of the evidence for each area of criticism based on the scale in Box 1.

Table 2. Breakdown of articles included in the review.

Number of studies included		105
Type of document	Peer-reviewed publication	70
	Other articles from journals, not peer reviewed (e.g. comment pieces)	22
	Grey literature	8
	Working paper	1
	Book chapter	4
Format of document	Commentary	21
	Review	15
	Empirical study	69
Type of data used (empirical studies)	Quantitative	53
	Qualitative	12
	Mixed methods	4
Subject focus	Biomedical	13
	Wider health	45
	Wider research	47
Quality of studies (GRADE)	0 (Lower quality, e.g. commentary with no or limited evidence base)	14
	1 (Triple-downgraded randomized trials, downgraded observational studies, or case series/case reports)	35
	2 (Double-downgraded randomized trials or observational studies)	38
	3 (Downgraded randomized trials or upgraded observational studies)	17
	4 (Randomized Trials or double-upgraded observational studies)	1

Box 1. Scale for quality of evidence

1. Assumptions: Intuitive assumptions and widely shared beliefs prevail
2. Suggestive: There is insufficient evidence to draw a clear conclusion (but the evidence is at least suggestive)
3. Conflicting: There are conflicting results from well-conducted studies
4. Agreement: A number of well-conducted studies agree
5. Compelling: Systematic reviews are compelling.

When synthesising our findings, we also drew on our previous review of the topic (Ismail *et al.*, 2009)

Results

We summarise our findings in Table 3 with each discussed in detail below.

Is peer review an effective system for awarding grants?

The meaning of 'best' science is not fixed. What constitutes the 'best' science will vary, however it may include research that is innovative, interdisciplinary and applied. This section considers biases against any particular type of research and whether peer review is a good predictor of future success.

Peer review is probably anti-innovation. Braben (2004) has suggested that supporting highly innovative research is important because it drives technological change and economic growth – an idea increasingly embraced by research funders. NIH has expressed concern at falling numbers of innovative or risky applications, suggesting 'competitive pressures have pushed researchers to submit more conservative applications' (Kaplan, 2005; Scarpa, 2006). Low success rates may have exacerbated the situation, inducing 'conservative, short-term thinking in applicants, reviewers, and funders' (Alberts *et al.*, 2014). On the other hand, a system is necessary to distinguish between innovative research and that grounded in 'reckless speculation' (Hackett & Chubin, 2003). Although 'innovative research' and 'high-risk research' are often conflated, they are not necessarily synonymous, here we include both aspects of innovation.

Innovative proposals may have less preceding work supporting them, and hence receive less praise from reviewers (RIN, 2010; Spier, 2002). This lack of preceding work requires less risk-averse mind-set from the reviewer (Spier, 2002). Innovative proposals from young researchers may suffer a 'double disadvantage': lacking previous work, both because of their novelty and the researcher's shorter track record.

The challenge of supporting innovation is not new, in 1977, Thomas Kuhn wrote of an 'essential tension' between originality and tradition. These tensions were also included in a 2006 UK Treasury report which noted 'the UK is still susceptible to a charge of risk aversion, as classic peer review criteria emphasise tests of scholarship over potential impact' (Treasury, 2006, p. 16). Empirical evidence of this problem come from recent work identifying lower scoring of novel proposals, even controlling for factors such as proposal quality, further this deficit could not be explained by the novel proposals being less feasible (Boudreau *et al.*, 2012; Boudreau *et al.*, 2016).

Risk aversion may also affect the preparation of applications: Fang & Casadevall (2009) suggested that falling success rates lead

[1]GRADE is an internationally accepted system for the assessment of evidence quality. GRADE offers four levels of evidence quality: high, moderate, low, and very low. Randomised trials begin as high-quality evidence and observational studies as low-quality evidence, and studies may be downgraded as a result of limitations in study design or implementation, imprecision of estimates, variability in results, indirectness of evidence, or publication bias. Equally, quality may be upgraded based on a very large magnitude of effect or if all plausible biases would reduce an apparent effect (Guyatt *et al.*, 2008).

Table 3. Summary of evidence from the literature regarding the effectiveness and burden of peer review.

Evaluation question	General critique	Particular criticism(s)	Is the criticism valid?	Strength of the evidence base
Is peer review an effective system for awarding grants?	Peer review does not fund the best science	It is anti-innovation	Yes	Suggestive
		It does not reward interdisciplinary work	Unclear	Suggestive
		It does not reward translational/ applied research	Unclear	Suggestive
		It is only a weak predictor of future performance	Yes	Agreement
	Peer review is unreliable	Ratings vary considerably between reviewers	Yes	Agreement
		It struggles to achieve an acceptable level of consistency	Unclear	Conflicting
	Peer review is unfair	It is gender-biased	Unclear	Conflicting
		It is age-biased	Unclear	Conflicting
		It is biased by cognitive particularism	Unclear	Conflicting
		It is open to cronyism	Yes	Agreement
	Peer review is not accountable	Review anonymity reduces transparency	N/A	N/A
	Peer review is not timely	It slows down the grant award process detrimentally	Unclear	Suggestive
	Peer review does not have the confidence of key stakeholders	It is not the preferred method of resource allocation	No	Agreement
What is the burden of peer review on the research system?	Peer review is an overly burdensome way of distributing research funding	Burden of peer review is increasing	Yes	Agreement
		Burden of the peer review system is high and falls primarily on the applicants	Yes	Agreement

to conservatism because of the perceived increased risk associated with innovative proposals.

Approaches to these problems include using reviewers with different cognitive biases for different schemes – specifically targeting specialists in translational or high-risk, innovative research (Langfeldt, 2001). This approach has been used (though not evaluated) in NIH's high-risk, high-reward Pioneer awards (Gewin, 2012).

Making 'innovation' an assessment criteria is another approach (Lindner *et al.*, 2016; Luukkonen, 2012). Views on this are mixed, some suggesting panels lack the expertise to assess innovation (Costello, 2010), whilst others see the approach as effective (Spiegel, 2010). Analysis of NIH application scores suggests that those for innovation are closely correlated with overall scores (Lindner *et al.*, 2016).

Other analysis (Giraudeau *et al.*, 2011; Linton, 2016), suggests that disagreement among scoring could be used to identify innovative research – high disagreement being taken as an indicator of work with high potential but also high risk. Similarly, Lee (2015) suggests combating conservatism by increasing the weight given to criteria – such as innovation – which are typically underweighted by reviewers.

An approach that sidesteps the issue is to select researchers purely on their merit, regardless of the research they plan to conduct. Researchers then have freedom to pursue new and novel ideas and work flexibly, as opportunities arise (e.g. by the MacArthur Fellows programme[4]).

Finally, Holliday & Robotin (2010) suggest that a Delphi process (a structured deliberative process) could be used to assess the merits of research 'in situations where the available scientific evidence is limited and if review panels have widely divergent opinions'. The process was also found to be efficient and flexible from a time perspective.

[4]As of 5 January 2017: https://www.macfound.org/programs/fellows/

It is not clear if peer review treats interdisciplinary research fairly. Critics argue interdisciplinary research is disadvantaged because (1) interdisciplinary proposal reviews may have to combine multiple distinct understandings of 'quality' – undermining the strength of the review (Feller, 2006), and (2) it is more difficult to identify 'peers' to review such work. This latter challenge is exacerbated by the standard structure of peer review processes in which only a few reviewers examine each proposal in detail, or at the initial stages, reducing the breadth of reviewing expertise further (Gluckman, 2012).

A study on the US National Science Foundation (NSF) revealed that, in interdisciplinary studies at least, peer review favours 'research that is performed by academics, in the sciences, and that falls completely within the reviewers' own domain of expertise' (Porter & Rossini, 1985, p. 37). With interdisciplinary teams it can be hard to isolate the contribution of each researcher, which can reduce the investigators chance of getting further funding by 'weakening' their track record (Cooksey, 2006a)

There has been limited further work in this area since 2009. Increasing the size of the review panel and broadening the range of expertise and disciplines present has been suggested as a way to address these problems. However, this increases burden and can only work if the role of the initial in-depth reviewer(s) is diminished (Gluckman, 2012).

It is not clear if peer review fairly assesses applied research. The Cooksey Report on health research funding in the UK noted that peer review 'can in some instances inhibit programmes in translational and applied health research' (Cooksey, 2006b). The report suggested that one reason for this inhibition was because peer review prevented the iterative development of research projects where funder and researcher worked together. Cooksey also suggested that because applied researchers publish in specialist (i.e. lower-impact) journals, they received less credit for publications than basic researchers. Including research users and considering the likely impact of research as part of the funding process may address these concerns. In our 2009 review, we noted the Canadian Health Services Research Foundation pioneering work through the use of 'merit review panels' to evaluate proposals, combining members from both academic and wider user/policy communities. This approach has now spread to other major funders, notably NIHR. Considering impact at the application stage – an approach criticised for disadvantaging innovative research – is likely to be beneficial when reviewing research closer to application.

The evidence around peer review's bias against applied research is not strong and has changed little since 2009. It is hard to know what criteria individual reviewers apply, as studies are hampered by methodological problems and funders are reluctant to release scores from peer review panels (Feller, 2006). While several studies have examined how reviewers assess proposals in the humanities and social sciences (Guetzkow *et al.*, 2004; Mansilla, 2006), work in the natural sciences is lacking. A study of NIH shows the success rate of clinical research proposals is marginally less than those for laboratory research (Kotchen *et al.*, 2004). This is in line with a recent CIHR study showing that health services and policy research applications were less successful than biomedical research applications (Tamblyn *et al.*, 2016).

Peer review is at best only a weak predictor of future performance. Work by Fang & Casadevall suggests peer review can 'winnow' out bad research proposals (Fang & Casadevall, 2012). However, recent studies from several NIH Institutes and the Netherlands have challenged the idea that peer review can effectively select the best research. Studies comparing percentile application rankings with the research's subsequent bibliometric performance found no association (Danthi *et al.*, 2014; Danthi *et al.*, 2015; Doyle *et al.*, 2015; Fang *et al.*, 2016; Kaltman *et al.*, 2014; van den Besselaar & Sandström, 2015). Two further such studies found that grant review outcomes only weakly predict bibliometric performance (Lauer *et al.*, 2015; Reinhart, 2009). Bibliometric analyses are by no means perfect measures of performance – only capturing a proxy of academic performance (Belter, 2015). Nonetheless, the findings suggest that peer review assessment is, at best, a crude predictor of performance.

Using an alternative metric, Galbraith *et al.* showed that peer reviewers' opinions were only weakly predictive of the commercial success of early stage technologies in small businesses (Galbraith *et al.*, 2010).

Fang & Casadevall (2012) comment that, while reviewers can usually identify the top 20–30 per cent of grant applications, going further to identify the top 10 per cent is 'impossible without a crystal ball or time machine' (p.898).

Is peer review reliable?

If peer review is reliable, the judgements of different peer reviewers on the same proposal should be highly correlated. The grounds for the continuing use of peer review would be severely undermined if systematic unreliability were demonstrated. Funders have been criticised for not making sufficient efforts to measure and monitor the reliability of assessments across reviewers (Fang & Casadevall, 2009). In this section, we consider two concerns surrounding peer review, namely individual reviews and overall consistency of decisionmaking – and how they might be addressed.

It is clear that ratings vary considerably between reviewers. *Single*-rater reliabilities[5] are not encouraging, but have been hampered by the methodological difficulties of modelling the complex interactions between reviewers in multi-stage peer review processes. In particular, the work of Jayasinghe *et al.* (2003) demonstrates a single-rater reliability correlation of just 0.21 for the humanities and social sciences, and an even lower correlation of 0.19 for the sciences. Similarly, Fogelholm *et al.* found an inter-rater reliability of around 0.23 for medical research (Fogelholm *et al.*, 2012). In contrast, two studies have found a higher level of agreement between reviwers. The first study which built in some of the complexities of the peer review process, found a dependent reliability[6] rating for individual peer reviewers of 0.80. The

[5]Defined as 'the correlation between two independent assessors of the same submissions across a large number of different submissions' (Jayasinghe *et al.*, 2003, p.280).

[6]In a multi-stage review process, the assessor at each evaluation stage will know the score given to a particular research proposal at the previous stage. This particular study assessed the reliability of grant peer review processes by determining the proportion of those applications for which the dependent ratings on the same proposal did not change from the first to the second and third stage.

second study on the review process for Marie Curie Actions (a major EU funding stream) measured inter-rater reliability based on the average deviation in scores between raters, and found a high level of agreement (Pina *et al.*, 2015).

Strikingly, the chance of improvements from initial ratings during panel discussion is virtually nil (e.g. from 'no award' or 'possible award' to 'award'). This suggests that initial triage of applications may be preferable to re-rating rounds (Bornmann, *et al.*).

Increasing diversity of background and discipline of peer reviewers also reduces rating consistency. Lobb *et al.* (2013) identified a low intra-class correlation coefficient (0.12) when comparing reviewers from a research, practice or policy background. They also noted that the level of agreement among experts from different disciplines was considerably lower than that among adjudicators of the same discipline, meaning that the presence of several practitioners from the same discipline area could have the potential to skew funding outcomes, depending on the wider makeup of the panel. This suggests that peer review processes may not work well for transdisciplinary teams integrating both academic and non-academic experts. Taking a different perspective, Reinhart found that although the global intra-class correlation coefficient was 0.41, there were considerable differences between fields, for example, biology (0.45) versus medicine (0.20) (Reinhart, 2009).

There is conflicting evidence on whether peer review can achieve acceptable levels of decisionmaking consistency. Existing studies offer mixed judgements on the reliability of grant peer review. Bornmann identified a threshold of 80–90 per cent as the expectation for agreement for this kind of decisionmaking (Bornmann *et al.*, 2008). Two early studies (Cole *et al.*, 1981; Hodgson, 1997) we noted in 2009 found reliability rates across funding boards of 75 and 73 per cent respectively for funding decisions which they felt was a satisfactory level of agreement. More recent evidence is mixed. The most recent study comparing the outcome of two independent panels found an agreement rate of 83 per cent (Clarke *et al.*, 2016), whilst a previous study in 2012 was less favourable, showing agreement levels of 65–69 per cent (Fogelholm *et al.*, 2012).

Graves *et al.* (2011) examined the variability of panel members' individual scores and calculated how this translates into the variability of overall proposal ranking, and hence funding decisions. They found that such variability could affect the outcome for 29 per cent of the proposals considered, and that variability differed widely between panels. Abdoul *et al.* have suggested that scoring variability might be partially explained by differences in reviewer behaviour, such as the time taken to do the assessment, assessment methods, and variation in the relative weighting of different criteria by different reviewers (Abdoul *et al.*, 2012).

Recent studies focusing more on the impact of panel meetings have shown very limited effects on improving consistency and reliability. Fogelholm *et al.* (2012) suggested that mean reviewer scores prior to the panel meeting were similar to the panel consensus score. The authors concluded that using the mean reviewers' scores was a practical and economical alternative. Similarly, although Pina *et al.* (2015) identified both a subset of panels and subset of proposals with high levels of disagreement, where consensus meetings improved agreement, across the whole population they could not detect an overall improvement in agreement.

In contrast, Martin *et al.* (2010) found meeting discussions had an important effect in more than 13 per cent of applications in their analysis of a sample of standard (R01) NIH research grant applications.

Two funders have experimented with, and evaluated, virtual peer review both by teleconference and through the use of Second Life, a virtual world. NIH estimated that using Second Life telepresence, peer review could cut panel costs by one third (Bohannon, 2011). Pier *et al.* (2015) compared videoconference and face-to-face panels. They set up one videoconference and three face-to-face panels modelled on NIH review procedures, concluding that scoring was similar between face-to-face and videoconference panels. Both the Bohannon and Pier studies of virtual panels noted that participants valued the social aspects of meeting in person and preferred the face-to-face arrangements.

Gallo *et al.* (2013) examined four years of peer review discussions, two years face-to-face and two years teleconferencing. They found minimal differences in merit score distribution, inter-rater reliability or reviewer demographics. They also noted that panel discussion, of any type, only affects the funding decision for around 10 per cent of applications relative to original scores.

Approaches to improve reliability have been tried. The NIH peer review self-study suggested some possible improvements to the peer review process to combat low reliability, focusing principally on better training for reviewers (NIH, 2008). NIH suggested such training should focus on: (1) emphasising the strengths (rather than weaknesses) of research proposals; (2) focusing on the potential impact of research; (3) reviewing the merit of the proposal and not re-writing it; (4) recognising the problem of implicit bias in study sections; (5) using benchmark applications during panel meetings to provide review guidelines; and (6) pointing out potential bias towards lesser known applicant organisations.

Recent work by Sattler *et al.* (2015) has evaluated the effect this type of brief training programme. The study found inter-rater reliability increased from 0.61 to 0.89, and the amount of time spent reviewing also increased, for both new and experienced reviewers.

If inconsistency stems from discrepancies in review quality (which is by no means clear), it might be feasible to evaluate the quality of reviews, although this approach has its own challenges – for example, what is a 'good' review? If a review is not consistent with other review does that intrinsically make it 'bad'? It could be the outlier picking up on the true potential of an innovative application. However, this approach is used by many funders, as shown in a report by the European Science Foundation (2011) which found in a survey of European research funders that more than half (60 per cent) evaluate the quality of all reviews as standard practice using a range of criteria (e.g. completeness, level of substantiation, appropriateness, comprehensibility, timeliness and usefulness), and

may return the review to a reviewer or reject the review. Organisations felt that review quality was higher where these checks were made, but noted little difference quality between cases where all reviews are evaluated versus just a sample. However, no data was available to assess these suggestions, and no empirical analysis had been carried out. Adding such an evaluation process clearly adds to the burden of the process.

Is peer review fair? Having considered the evidence suggesting that consensus on peer review decisions is rare, what factors might underlie the observed discrepancies? To what extent is peer review open to the same allegations of bias that plague science more widely, particularly around gender, race, intellectual school or institutional affiliation? A recent study (Day, 2015) has shown that low levels of passive bias as well as individual cases of significant active bias among reviewers can have significant impacts on the outcomes of a grant peer review process. In this section we consider the potential for bias in peer review across four main areas: gender, age, cronyism and cognitive particularism.

Bias could occur at various places in the peer review process. While bias on the part of the peer reviewers themselves (such as sexism or racism) has received considerable attention in the literature, funding competitions can be biased through eligibility and award selection criteria. Such criteria may be prejudiced against early career researchers or innovative research – although there is no strong evidence that this occurs. In addition wider systemic biases may mean that the number of applications received is lower from particular groups.

Blinding of applications provides a defence against the most obvious abuses by reviewers – rejecting proposals on the grounds of race, gender, institutional affiliation and so forth (Lee *et al.*, 2012). A study from South Korea by Lee *et al.* (2000) demonstrated a significant bias in sighted proposal evaluation towards those from particular research departments, senior researchers, and those already academically recognised. This is reinforced by a review of studies by the NSF, which found only 'a weak correlation' between panel ratings of blinded short version and unblinded full versions of the same applications (Bhattacharjee, 2012). While some funding bodies now routinely attempt to anonymise proposals *before* passing them on to reviewers, there is some dispute as to whether anonymisation is truly possible. Some authors contend that some degree of identification is always possible from anonymised research proposals (Bhattacharjee, 2012)

There is a substantial body of conflicting evidence on whether peer review is gender biased. The evidence on gender bias is inconclusive. Studies suggesting bias include an important study of the grant peer review system of the Swedish Medical Research Council strongly suggested that reviewers were unable to judge scientific merit independently of gender (Wenneras & Wold, 1997). These findings were supported by a subsequent meta-analysis of 21 studies on this topic, which found that grant applications submitted by men were 7 per cent more likely to be approved than those submitted by women (Bornmann *et al.*, 2007).[7] Furthermore, recent studies have also found evidence of gender bias (Jang *et al.*, 2016; Kaatz *et al.*, 2014; Kaatz *et al.*, 2015; Tamblyn *et al.*, 2016; van der Lee & Ellemers, 2015; Volker & Steenbeek, 2015). For example, van der Lee & Ellemers (2015) reported a 4 per cent 'loss' of women during the grant review process for awards to early career scientists by the Netherlands Organization for Scientific Research (NWO). In a review of research on gender bias by Kaatz *et al.* (2014), women generally have lower rates of publication and lower success rates for high-status research awards than do men.

On the other hand, a review of the gender bias literature by Ceci & Williams (2011) showed that the weight of evidence suggests that peer review is fair across gender, with all smaller-scale studies analysed, along with all but one of the large-scale studies, failing to replicate Wenneras & Wold's findings. And even for the remaining large-scale study the findings were reversed by a reanalysis. The lack of gender bias has been supported by several subsequent studies (Marsh *et al.*, 2011; Mutz *et al.*, 2012; Reinhart, 2009; Turner *et al.*, 2014; Van Arensbergen *et al.*, 2013).

There is a small conflicting evidence base on whether peer review is biased by age. Although review processes that partly rely on the previous publications or funding successes of the applicant may be biased against early career researchers, Jayasinghe *et al.* (2001); Jayasinghe *et al.* (2003) found that the age of the applicants did not directly impact upon grant success, a findings supported by Reinhart (2009). However, this finding is directly contradicted by a study comparing sighted and blinded reviews of research grant proposals in South Korea (Lee *et al.*, 2000). A subsequent study, also based in South Korea (Jang *et al.*, 2016), found that evaluation scores and selection success rates decline with age.

Concerns about age bias are closely tied to concerns about bias against early career researchers, who may be disadvantaged through lacking preliminary results or a substantial portfolio of work. The challenges of providing adequate support for early career researchers is widely recognised (Bazeley, 2003) and was raised in a 2008 NIH review which identified significant decreases early career success rates which could not be accounted for by variations in application quality (NIH, 2008). Similar concerns were also noted by Spiegel (2010), who showed that the average age researchers won their first full NIH project grant awards (R01) had been steadily increasing. Since then, the NIH has introduced measures aimed at equalising success rates for new and established investigators for new (not renewal) applications.

[7]Bornmann *et al.* (2007) are clear, however, that the reasons for this observed discrepancy are not known. This is important because aggregation effects over a range of fields of study *may* – as the authors acknowledge –create strong statistical effects implying gender bias. The authors also suggest that future improvements to the model will need to take into account the cohort of application, since the study described here covered publications produced over the period 1979–2004, and there have been significant changes to reduce gender bias in science and science funding over this period.

There is evidence that peer review suffers from cronyism. Cronyism is a concern for many major funders, who have detailed conflict of interest processes in place to counter the presence or perception of such biases. However, (Wenneras & Wold, 1997) show that prior affiliation with a reviewer considerably increased a researcher's chances of funding, Similarly, a large-scale study of applications to the National Science Foundation of Korea found that applications reviewed by previous or current affiliates were more likely to be successful (Jang *et al.*, 2016). A review of NSF proposals by (Bhattacharjee, 2012), found that panel assessments of full proposals and shorter anonymised versions of the same proposals showed weak correlations.

Luukkonen (2012) notes that panel debate may fail to counter crude forms of cronyism since panels often cover a wide area of research, and each specific area is only represented by a few experts, so the other members may defer to the experts' knowledge. Members of funding panels may also benefit directly from their membership. One study noted that panel members submit more applications, and have more grant awards (van den Besselaar, 2012). The challenge in this area is separating factors such as good researchers who submit more applications being selected to join panels or having a better sense of what makes a good application, from nepotism.

There is conflicting evidence on whether peer review demonstrates cognitive particularism (favouring your own field or way of thinking). The idea that reviewers and panel members will favour proposals in their own fields or that align with their ways of thinking has been termed 'cognitive particularism' (Travis & Collins, 1991). Fang & Casadevall (2009), suggest that 'reviewer biases favour topics well understood and appreciated by the [funding panel]' (p.930). Travis & Collins (1991) found that reviewers tend to favour proposals supporting their own school of thought, and argues that this is likely to have a much bigger impact on the direction of science than institutional bias or cronyism identified by other studies (Langfeldt, 2006; Wenneras & Wold, 1997). Research by Li (2015) suggests the same. Work by Wang & Sandström (2015) suggests that 'cognitive distance' may influence reviewer decisions in a more complicated way, with reviewers more likely to favour applications in areas they are either very familiar with, or completely unfamiliar with. Other studies find that reviewers are more critical of applications in areas of their own expertise (Boudreau *et al.*, 2016; Gallo *et al.*, 2016).

A number of studies suggest that studies in molecular biology are more likely to be successful in comparison to other fields of bioscience. (Bornmann & Daniel, 2006) found a slight statistical effect and further studies reveal that peer-reviewed grant proposals in molecular biology tend to have a better chance of receiving grant funding than proposals in other bioscience fields (Kotchen *et al.*, 2004; Taylor, 2001).

There is also dispute about how to resolve this potential problem Alberts *et al.* (2014) suggests that such effects could be countered by broadening 'the range of scientific problems judged by each group and include[ing] a diversity of fields on each panel', suggesting that 'senior scientists with a wide appreciation for different fields can play important roles by counteracting the tendency of specialists to overvalue work in their own field' (p.5777). However, Li (2015) advises caution, noting that though evaluators may be biased in favour of projects in their own area, they are also likely to be better able to assess the quality of those projects, and the benefits of this expertise may well outweigh any possible biases.

Is peer review timely?

There is suggestive evidence that the peer review process slows, and hinders, the progress of research. In some cases such as an emerging epidemic the time taken by peer review could reduce the number of people benefiting from the research, such slowing of the research process could also reduce the economic viability of a new product, (e.g. Agres, 2005; Cures, 2005; Daniels, 2004; Roy, 1985). The many stages of grant peer review can take from 9 to 18 months from submission to funding. It is less clear how often this time significantly hinders the progress of science. In the health sciences, research is one of many steps in develop new treatments and practices (Hanney *et al.*, 2015). Research suggests that the time required for translation of research from initial idea to adopted practice is around 17 years, so peer review may be a relatively small contributor, however any one translation pathway may have multiple stages of peer review (Morris *et al.*, 2011).

There is good evidence that peer review has the support of most major scientific stakeholders. Though criticism of the peer review process abounds, empirical evidence, though limited, indicates that support for peer review amongst the academic community remain strong (Bornmann, 2011; Wooding & Grant, 2003). The dominance of peer review across funding systems internationally suggests it has the confidence of institutional stakeholders. A recent review of literature about the NIH peer review processes found a firm belief in the transparency and objectivity of peer review amongst grant reviewers (Miner, 2011).

In contrast an emerging body of literature suggests traditional academic peer review may not be appropriate for all types of research. A recent study on indigenous research showed the competitive nature of peer review was counterproductive and that peer review did not have the confidence of relevant stakeholders (Street *et al.*, 2009). Similar concerns have been expressed about the assessment of community engagement proposals (Ahmed & Palermo, 2010).

What is the burden of peer review on the research system?

The burden of peer review is increasing. In a survey of 28 biomedical research funding organisations across 19 countries (Schroter *et al.*, 2010), declined review requests, late reports and administrative burden were the most frequently mentioned challenges, and all organisations reported an increase in burden in the previous five years (although they reported that the quality of reviews had remained the same). A study by the Royal Society of New Zealand reported a similar increase in the difficulty of recruiting senior reviewers (Gluckman, 2012).

The burden of the peer review system is high and falls primarily on the applicants. The overall monetised cost of the peer review system, including application preparation, has been estimated to account for as much as 20–35 per cent of the allocated budget (Gluckman, 2012). Graves *et al.* (2011) report that the monetised costs of the application system for NHMRC are $14,000 per grant,

whilst extrapolating RCUK (Research Councils UK, 2006) estimates suggests that the costs of the application process are 10–17 per cent of the total cost of research. An evaluation of the CIHR Operating Open Grants Program (OOGP) found the application cost of OOGP grants to be Can$14,000 (Peckham *et al.*, 2012). When providing congressional testimony individual researchers have estimated that as much as 60 per cent of their time is devoted to seeking funding (Fang & Casadevall, 2009).

Burden on applicants. The bulk of the resources consumed by the peer review process are in the writing and reviewing of applications. RCUK work showed the distribution of monetised burden was 74 per cent in application production, 21 per cent in reviewing process (including time of reviewers, panel membership and modifying proposals), and 5 per cent in Research Council costs and payments to reviewers (Research Councils UK, 2006).

More recent work by Graves *et al.* (2011) used a small survey of NHMRC researchers to estimate that the burden fell even more heavily on the applicants, assigning a split of 85 per cent for application production, 9 per cent for reviewing and 5 per cent for administration. Barnett *et al.* (2015) reinforced this conclusion with a larger survey of 285 applicants who had submitted 632 proposals to four health services research funding rounds from May 2012 to November 2013, at the Australian Centre for Health Services Innovation. A review by the New Zealand Royal Society made a similar estimate of the burden shouldered by the applicants – pegging it at 80 per cent (Gluckman, 2012).

In contrast two studies of the Natural Sciences and Engineering Research Council (NSERC) of Canada peer review process and came to strikingly different conclusions. Gordon & Poulin (2009) estimated the cost of the NSERC system, including application preparation, review and administration costs at Can$44m. They suggest this money could alternatively provide all researchers in the field with an annual baseline grant of Can$30,000. However, Roorda (2009) takes issue with Gordon and Poulin's assumptions suggesting they have overestimated costs by a factor of 23. The correct answer appears to be in between – there is disagreement about how the costs should be allocated and neither side provides a justification of their estimates of the time spent on grant preparation (the key driver).

Herbert *et al.* (2013) suggest burden on NHMRC applicants could be reduced by simplifying the application process (currently 80–120 page applications). Other examples of funding agencies reducing the length and complexity of applications include NIH did cut the length of their applications for R01s[8] from 25 pages to 12 in 2009, although there were calls to make the application even shorter (Fang & Casadevall, 2009).

[8]The Research Project Grant (R01) is the original and historically oldest grant mechanism used by NIH. The R01 provides support for health-related research and development based on the mission of the NIH. R01s can be investigator-initiated or can solicited via a Request for Applications.

Barnett *et al.* (2015) examined the effect of reducing the complexity of the application. Surprisingly, they found that reducing application complexity slightly increased preparation time. They suggest that this may be because researchers allocate a fixed fraction of their time to application preparation. Theoretical work by Geard & Noble (2010) using agent based modelling found that applicants devote 'excessive' time to proposal preparation (Geard & Noble, 2010). Barnett *et al.* (2015) examined four rounds of a funding scheme in Australian which significantly shortened the application (to 1,200 words). Qualitative feedback was positive, suggesting it took seven days to develop an application, but generalisability is limited. The level of effort devoted to application preparation is all the more striking given Herbert *et al.*'s (2013) finding that increased effort did not translate into increased success rates.

A few qualitative studies have examined the burden of the system on particular groups of researchers and the wider implications on researchers' quality of life. A survey of 215 NHMRC applicants concluded that the 'impact of preparing grant proposals for a single annual deadline is stressful, time consuming and conflicts with family responsibilities' (p.1), although it did not quantify the effects or time taken (Herbert *et al.*, 2014). A study of early career investigators applying for funding at CIHR identified the application process as burdensome and noted the decrease in success rates for open operating grants from 30 per cent in 2005–2006 to 15 per cent in 2014–2015 (Association of Canadian Early Career Health Researchers, 2016).

The institutional costs of application preparation were examined by the US Government Accountability Office (GAO) in 2016, which concluded that pre-award requirements for applicants to develop and submit detailed documentation for grant proposals, and increased prescriptiveness of certain requirements, had increased universities' workload and costs, but the study (GAO, 2016) did not quantify these increases

Burden on reviewers and panel members. Time invested by reviewers and panel members is consistently identified as the second-highest monetised cost of peer review, making up about 15 per cent of the burden. Two types of studies carried out in this area have both aimed at optimising the process, balancing the trade-off between burden and quality to achieve efficiency.

The first study approach trialled simplified processes for grant review to test how much time they save and whether they affected funding decisions (Herbert *et al.*, 2015) – particular the use of a shortened application form and smaller review panels. They found the simplified processes achieved agreement with the current award system of close to 75 per cent (which they suggested was the 'acceptable' threshold based on a review of previous surveys), at estimated savings of 33–78 per cent of review costs.

The second study used statistical techniques to estimate the optimum number of reviewers (Snell, 2015) trading off improved reproducibility with additional reviewer burden. They found that

five reviewers were optimal; similar work by Graves *et al.* (2011) on a different funding scheme found 11 reviewers was the most effective number.

In addition to experimental changes there are examples of funding agency policy changes that have been examined. The NSF changed its review procedures in 2012 to reduce burden by introducing triage on short preliminary applications with a 75 per cent cull rate, with annual rather than six-monthly applications. The General Accountability Office has praised the system and it reduces administrative burden on programme officers. However, because several changes happened simultaneously, it is not clear whether this is because of the triaging. It also resulted in reduced success rates, partly because of more applications (perhaps because they were easier to write but also because of funding reductions (Mervis, 2016).

One of the drivers of the burden on funders is identifying appropriate reviewers for each proposal. Mervis (2014) reports on a radical experiment at NSF where applicants reviewed each other's grants (each applicant completing seven reviews), consequently reducing this burden to zero. To guard against applicants marking their competitors down, they were rewarded for scores that aligned with the other reviewers. The pilot allowed the number of reviews per proposal to be increased from three or four to seven and the reviews provided were more detailed. Because of the additional reviews, NSF was able to dispense with panel discussion, thus saving administrative costs.

Discussion

In this section we summarise our findings: firstly, on the availability of evidence, considering the scope and coverage of the existing literature; secondly, on what that the evidence shows, and finally, highlighting the implications for health research funders.

Availability of evidence

Questions about the effectiveness and burden of peer review can be addressed at two levels. At a high level, does peer review support valuable science? And at a lower level, can the design of peer review systems be improved to increase effectiveness and reduce burden?

It is clear that the current system of funding has produced significant benefits for society, suggesting that peer review supports valuable science. However, whether peer review is demonstrably better than any other system is impossible to judge with certainty because of the lack of comparators: no funding agencies have made significant use of alternative systems.

Moving to the lower level, considering comparisons between or research on peer review systems, there is only a very small number of robust, well-conducted studies. Much of the literature identified is anecdotal in nature and we found no systematic reviews, underlining the fragility of the evidence base. However, we did identify a series of robust, high-quality studies that have been carried out since our last review in 2009. Despite this new work it is still true that most studies examine the peer review process of one particular funder in one particular context, rather than looking across funders or contexts, and few go beyond process measures to judge effectiveness.

This persistent lack of evidence about the allocation of the 'inputs' to research is all the more striking given the advances in understanding the outputs and outcomes of research through research impact assessment over the last decade.

Findings from the available evidence

The central problem when assessing peer review is the lack of an absolute standard or 'ground truth' to judge against. There will be uncertainty in all peer review decisions - it is, after all, predicting the future. And there is evidence suggesting it is not a particularly good predictor, at least for bibliometric performance. At present most funders do not capture, use, or even acknowledge this uncertainty, despite clear evidence of inconsistency in peer review ratings and mixed evidence on the reproducibility of panel decisions.

These is good evidence that peer review suffers from biases. The strongest evidence is of a bias against innovation and although a range of improvements have been suggested, none have been robustly evaluated. There is some evidence peer review is influenced by cognitive distance and suffers from cronyism and suggestive evidence that there are age biases. Considerable work has been done on gender bias, with conflicting results, which illustrates the challenges of accounting for biases outside the scope of the peer review process, for example through eligibility or the culture of the wider scientific system.

Though the problem of burden is widely recognised, funders' considerations often focus on their own and reviewers' burden as these are more immediately visible (and costly) to them. However, it is clear that the burden largely falls on applicants (rather than reviewers or panel members).

Falling success rates across many funders compound the burden on applicants. One way to address these challenges could be to reduce the complexity of the application process, with evidence suggesting similar decisions can be made with much shorter applications and less information. However, small decreases in application length do not seem to translate into application preparation time so such changes would need to be carefully evaluated.

Despite the plethora of comment pieces criticising the peer review system, there is no empirical evidence suggesting whether peer review has more or less support among key stakeholders than it did in 2009.

Potential improvements

Improving effectiveness. This section outlines our reflections on ideas for improving peer review processes.

We feel the uncertainty in peer review - clear in the inconsistency of ratings and weak predictive power in terms of future academic performance - should be acknowledged, captured and used to improve decision making and for analysis. Reviewers should be asked both for their rating of the proposal and a measure of their confidence in this rating - some smaller funders, such as the Villum and Velux Foundations in Denmark, are starting to implement such systems. Funders could also analyse levels of disagreement between

reviewers, which may be an indicator of innovative research (Linton, 2016), or take a portfolio approach selecting projects scoring highly across different criteria, including innovation (Lee, 2015).

A second approach is to acknowledge the difficulty of predicting the future and introduce an explicit element of randomness into the allocation system. This could be done to differing extents – from completely random allocation of funding to the use of a lottery system within set groups of applicants. Fang & Casadevall (2016) propose a two-stage system, in which the best applications are identified and then a smaller percentage are funded using a lottery. Avin (2015) proposes using two thresholds, above the higher threshold all applications are funded and below the lower threshold all applications are rejected, applications between the two thresholds are funded at random, effectively blurring the funding line.

A lottery approach should reduce biases in decision making since the selection from the fundable pool is random; however, applicant eligibility restrictions/selection for the lottery could reintroduce bias. Selecting into a fundable pool requires less fine-grained decisions addressing concerns about the reliability of peer review. The use of lottery systems is a promising, but politically challenging idea, so far is has only been used in very limited cases, such as the Explorer Grants offered by the Health Research Council of New Zealand[9], and as such we think it merits further empirical research (Barnett, 2016).

Other approaches to address bias include blinding of reviewers (e.g. Lee *et al.*, 2012), though the feasibility of this is debated ((Bhattacharjee, 2012). Other funders have also used training approaches to address bias (e.g. CIHR) and to improve quality of reviews (e.g. NIH, 2008) and there is limited evidence that the approach could reduce the discrepancies between reviewers (Sattler *et al.*, 2015).

Reducing Burden. Applicant burden should be considered as a priority compared to reviewer and administrative burden as it represents around 75% of the system burden. This can be addressed by reducing the level of burden or increasing the value unsuccessful applicants receive by applying. Changes to reduce burden need to be carefully evaluated as there is evidence that even significant reductions in application length/complexity may not reduce applicant burden as much as expected. An alternative approach is to make the process more valuable for the applicants. Reviewer and panel feedback may be one way to do this.

Technology provides ways to reduce the time burden of the peer review process for panel members and funders - for example by eliminating travel - and does not appear to significantly affect the outcomes. However, face-to-face discussion of applications brings other side-benefits, including social interaction and network formation, other research suggests these side-benefits may be important to the progress of science and hence may need to be supported in other ways if peer review is done remotely.

Improving the evidence base. It remains striking how little robust evidence is available about peer review as a method for grant allocation. Given the centrality of the peer review process in the current science funding system, there is a need for better evidence, not only on the overall effectiveness of peer review but also to help improve the design of peer review processes. We suggest three fruitful areas for investigator are the links between the peer review process and the wider context of science funding; the social processes of peer review and panel meetings.

System impacts affect the peer review process, and peer review changes affect the system, so both need to be considered together to understand the dynamic behaviour of the overall research process. All of the studies we identified considered aspects of the peer review system in isolation – for example tracking success rates or reviewer burden. However, system changes such as decreased funding, or changes in researcher demographics, often happen alongside changes to the peer review system.

Even in the fairly barren landscape of evidence we explored, it was startling that we could find no studies examining the social processes that occur during panel discussions – a central part of the peer review process. Such studies will clearly be challenging and require the cooperation of funders working in concert, but we feel are essential to understand how to optimise one of the fundamental processes of science.

At a more mundane level, funders should be more willing to experiment with, evaluate and publish results from evaluations of alternative approaches. Through our conversations with funders it appears that where analysis is carried out it is often not published, partly because of the extreme sensitivity around funding allocation procedures. Funders are not the only ones who need to take a more reflective approach: they will need the support of the wider scientific community to support such investigations, and acknowledge the lack of evidence about the primacy of the current system and the impossibility of achieving perfection.

Conclusions

Many criticisms of the peer review system reflect conflicts between the needs of stakeholders. Researchers look to peer review to uphold research standards and promote the 'best' science, while politicians and funders use it to provide accountability for spending (Viner *et al.*, 2004). This tension requires peer review to both protect the identities of reviewers while appearing transparent to applicants; to be innovative yet assure quality; to be based on human judgement yet free of human biases (Hackett & Chubin, 2003).

We think that current dissatisfaction with the peer review process is amplified by falling success rates, so it is important to remember

[9]As of 5 January 2017: http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposals/explorer-grants

that the concerns around peer review are heavily influenced by funding policy and the size of research budgets.

As a society, if we are to improve how we use our research funds, we need a better understanding of the peer review process. When making changes, funders should: build in before and after comparisons; strive to make data available for analysis; openly publish studies of their processes and work together on comparative analysis.

We need to overcome the reluctance of funders and scientists to acknowledge the uncertainties intrinsic to allocating research funding, and encourage them to experiment with peer review and other allocation processes.

Competing interests

No competing interests were disclosed.

Grant information

This report was produced with funding from the Canadian Institutes of Health Research.

Acknowledgements

The authors acknowledge the Performance and Accountability Branch at the Canadian Institutes of Health Research for funding this study, and Sarah Viehbeck, Shevaun Corey, Kwadwo Bosompra, Michael Goodyer and David Peckham in that Branch for their input and advice on the development of the work.

The views expressed in this report are those of the authors and do not necessarily reflect those of the Canadian Institutes of Health Research.

We would also like to thank our RAND Europe quality assurance reviewers Catherine Lichten and Gavin Cochrane for their helpful comments and suggestions.

References

Abdoul H, Perrey C, Amiel P, *et al.*: **Peer review of grant applications: criteria used and qualitative study of reviewer practices.** *PLoS One.* 2012; **7**(9): e46054.
PubMed Abstract | Publisher Full Text | Free Full Text

Agres T: **Venture capital, with a twist.** *Scientist.* 2005; 45–46.
Reference Source

Ahmed SM, Palermo AG: **Community engagement in research: frameworks for education and peer review.** *Am J Public Health.* 2010; **100**(8): 1380–1387.
PubMed Abstract | Publisher Full Text | Free Full Text

Alberts B, Kirschner MW, Tilghman S, *et al.*: **Rescuing US biomedical research from its systemic flaws.** *Proc Natl Acad Sci U S A.* 2014; **111**(16): 5773–5777.
PubMed Abstract | Publisher Full Text | Free Full Text

Association of Canadian Early Career Health Researchers: **Early Career Investigators (ECIs) in health research: final report of a cross-Canada survey.** 2016.
Reference Source

Avin S: **Funding Science by Lottery.** *European Studies in Philosophy of Science.* 2015; 111–126.
Publisher Full Text

Barnett AG: **Funding by Lottery: Political Problems and Research Opportunities.** *MBio.* 2016; **7**(4): pii: e01369-16.
PubMed Abstract | Publisher Full Text | Free Full Text

Barnett AG, Herbert DL, Campbell M, *et al.*: **Streamlined research funding using short proposals and accelerated peer review: an observational study.** *BMC Health Serv Res.* 2015; **15**: 55.
PubMed Abstract | Publisher Full Text | Free Full Text

Bazeley P: **Defining 'early career' in research.** *High Educ.* 2003; **45**(3): 257–279.
Publisher Full Text

Belter CW: **Bibliometric indicators: opportunities and limits.** *J Med Libr Assoc.* 2015; **103**(4): 219–221.
PubMed Abstract | Publisher Full Text | Free Full Text

Bhattacharjee Y: **Science funding. NSF's 'Big Pitch' tests anonymized grant reviews.** *Science.* 2012; **336**(6084): 969–970.
PubMed Abstract | Publisher Full Text

Bohannon J: **National Science Foundation. Meeting for peer review at a resort that's virtually free.** *Science.* 2011; **331**(6013): 27.
PubMed Abstract | Publisher Full Text

Bornmann L: **Scientific peer review.** *Inf Sci Technol.* 2011; **45**(1): 197–245.
Publisher Full Text

Bornmann L, Daniel HD: **Potential sources of bias in research fellowship assessments: effects of university prestige and field of study.** *Res Eval.* 2006; **15**(3): 209–219.
Publisher Full Text

Bornmann L, Mutz R, Daniel HD: **Latent Markov modeling applied to grant peer review.** *J Informetr.* 2008; **2**(3): 217–228.
Publisher Full Text

Bornmann L, Mutz R, Daniel HD: **Gender differences in grant peer review: a meta-analysis.** *J Informetr.* 2007; **1**(3): 226–238.
Publisher Full Text

Boudreau KJ, Guinan EC, Lakhani KR, *et al.*: **The Novelty Paradox & Bias for Normal Science: Evidence from Randomized Medical Grant Proposal Evaluations**. *Harvard Business School Working Paper.* 2012; 13–053.
Publisher Full Text

Boudreau KJ, Guinan EC, Lakhani KR, *et al.*: **Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science.** *Manage Sci.* 2016; **62**(10): 2765–2783.
PubMed Abstract | Publisher Full Text | Free Full Text

Braben D: **Pioneering Research: A Risk Worth Taking.** Hoboken, New Jersey: Wiley Interscience. 2004.
Reference Source

Ceci SJ, Williams WM: **Understanding current causes of women's underrepresentation in science.** *Proc Natl Acad Sci U S A.* 2011; **108**(8): 3157–3162.
PubMed Abstract | Publisher Full Text | Free Full Text

Clarke P, Herbert D, Graves N, *et al.*: **A randomized trial of fellowships for early career researchers finds a high reliability in funding decisions.** *J Clin Epidemiol.* 2016; **69**: 147–151.
PubMed Abstract | Publisher Full Text

Cole S, Cole JR, Simon GA: **Chance and consensus in peer review.** *Science.* 1981; **214**(4523): 881–886.
PubMed Abstract | Publisher Full Text

Cooksey D: **A review of UK health research funding**. 2006a.
Reference Source

Cooksey D: **A Review of UK Health Research Funding**. Norwich: HMSO. 2006b.
Reference Source

Costello LC: **Perspective: is NIH funding the "best science by the best scientists"? A critique of the NIH R01 research grant review policies.** *Acad Med.* 2010; **85**(5): 775–779.
PubMed Abstract | Publisher Full Text

Cures F: **'Investing in Innovation' Meeting Report.** 2005.

Daniels C: **The man who changed medicine.** *Fortune.* 2004; **150**(11): 90–4.
PubMed Abstract

Danthi N, Wu CO, Shi P, *et al.*: **Percentile ranking and citation impact of a large cohort of National Heart, Lung, and Blood Institute-funded cardiovascular R01 grants.** *Circ Res.* 2014; **114**(4): 600–606.
PubMed Abstract | Publisher Full Text | Free Full Text

Danthi NS, Wu CO, DiMichele DM, *et al.*: **Citation impact of NHLBI R01 grants**

funded through the American Recovery and Reinvestment Act as compared to R01 grants funded through a standard payline. *Circ Res.* 2015; **116**(5): 784–788.
PubMed Abstract | Publisher Full Text | Free Full Text

Day TE: **The big consequences of small biases: A simulation of peer review.** *Res Policy.* 2015; **44**(6): 1266–1270.
Publisher Full Text

Doyle JM, Quinn K, Bodenstein YA, *et al.*: **Association of percentile ranking with citation impact and productivity in a large cohort of *de novo* NIMH-funded R01 grants.** *Mol Psychiatry.* 2015; **20**(9): 1030–1036.
PubMed Abstract | Publisher Full Text

European Science Foundation: **ESF Survey Analysis Report on Peer Review Practices.** 2011.
Reference Source

Fang FC, Bowen A, Casadevall A: **NIH peer review percentile scores are poorly predictive of grant productivity.** *eLife.* 2016; **5**: pii: e13323.
PubMed Abstract | Publisher Full Text | Free Full Text

Fang FC, Casadevall A: **NIH peer review reform--change we need, or lipstick on a pig?** *Infect Immun.* 2009; **77**(3): 929–932.
PubMed Abstract | Publisher Full Text | Free Full Text

Fang FC, Casadevall A: **Reforming science: structural reforms.** *Infect Immun.* 2012; **80**(3): 897–901.
PubMed Abstract | Publisher Full Text | Free Full Text

Fang FC, Casadevall A: **Research Funding: the Case for a Modified Lottery.** *MBio.* 2016; **7**(2): e00422–00416.
PubMed Abstract | Publisher Full Text | Free Full Text

Feller I: **Multiple actors, multiple settings, multiple criteria: issues in assessing interdisciplinary research.** *Res Eval.* 2006; **15**(1): 5–16.
Publisher Full Text

Fogelholm M, Leppinen S, Auvinen A, *et al.*: **Panel discussion does not improve reliability of peer review for medical research grant proposals.** *J Clin Epidemiol.* 2012; **65**(1): 47–52.
PubMed Abstract | Publisher Full Text

Galbraith CS, DeNoble AF, Ehrlich SB, *et al.*: **Review panel consensus and post-decision commercial performance: a study of early stage technologies.** *J Technol Transfer.* 2010; **35**(2): 253–281.
Publisher Full Text

Gallo SA, Carpenter AS, Glisson SR: **Teleconference versus face-to-face scientific peer review of grant application: effects on review outcomes.** *PLoS One.* 2013; **8**(8): e71693.
PubMed Abstract | Publisher Full Text | Free Full Text

Gallo SA, Sullivan JH, Glisson SR: **The Influence of Peer Reviewer Expertise on the Evaluation of Research Funding Applications.** *PLoS One.* 2016; **11**(10): e0165147.
PubMed Abstract | Publisher Full Text | Free Full Text

GAO U: **Federal Research Grants: Opportunities Remain for Agencies to Streamline Administrative Requirements.** Washington: United States Government Accountability Office. 2016.
Reference Source

Geard N, Noble J: **Modelling academic research funding as a resource allocation problem**. *Paper presented at the 3rd World Congress on Social Simulation.* 2010.
Reference Source

Gewin V: **Risky research: The sky's the limit.** *Nature.* 2012; **487**(7407): 395–397.
PubMed Abstract | Publisher Full Text

Giraudeau B, Leyrat C, Le Gouge A, *et al.*: **Peer Review of Grant Applications: A Simple Method to Identify Proposals with Discordant Reviews.** *PLoS One.* 2011; **6**(11): e27557.
PubMed Abstract | Publisher Full Text | Free Full Text

Gluckman P: **Which science to fund: time to review peer review?** New Zealand Office of the Prime Minister's Science Advisory Committee. 2012.
Reference Source

Gordon R, Poulin BJ: **Cost of the NSERC Science Grant Peer Review System exceeds the cost of giving every qualified researcher a baseline grant.** *Account Res.* 2009; **16**(1): 13–40.
PubMed Abstract | Publisher Full Text

Graves N, Barnett AG, Clarke P: **Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel.** *BMJ.* 2011; **343**: d4797.
PubMed Abstract | Publisher Full Text | Free Full Text

Guetzkow J, Lamont M, Mallard G: **What is originality in the humanities and social sciences?** *Am Sociol Rev.* 2004; **69**(2): 190–212.
Publisher Full Text

Guyatt GH, Oxman AD, Vist GE, *et al.*: **GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.** *BMJ.* 2008; **336**(7650)**:** 924–6.
PubMed Abstract | Publisher Full Text | Free Full Text

Hackett E, Chubin D: **Peer review for the 21st century: Applications to education research**. *Paper presented at the National Research Council Workshop.* Washington, DC. 2003.
Reference Source

Hanney SR, Castle-Clarke S, Grant J, *et al.*: **How long does biomedical research take? Studying the time taken between biomedical and health research and its translation into products, policy, and practice.** *Health Res Policy Syst.* 2015; **13**(1): 1.
PubMed Abstract | Publisher Full Text | Free Full Text

Herbert DL, Barnett AG, Clarke P, *et al.*: **On the time spent preparing grant proposals: an observational study of Australian researchers.** *BMJ Open.* 2013; **3**(5): pii: e002800.
PubMed Abstract | Publisher Full Text | Free Full Text

Herbert DL, Coveney J, Clarke P, *et al.*: **The impact of funding deadlines on personal workloads, stress and family relationships: a qualitative study of Australian researchers.** *BMJ Open.* 2014; **4**(3): e004462.
PubMed Abstract | Publisher Full Text | Free Full Text

Herbert DL, Graves N, Clarke P, *et al.*: **Using simplified peer review processes to fund research: a prospective study.** *BMJ Open.* 2015; **5**(7): e008380.
PubMed Abstract | Publisher Full Text | Free Full Text

Hodgson C: **How reliable is peer review? An examination of operating grant proposals simultaneously submitted to two similar peer review systems.** *J Clin Epidemiol.* 1997; **50**(11): 1189–1195.
PubMed Abstract | Publisher Full Text

Holliday C, Robotin M: **The Delphi process: a solution for reviewing novel grant applications.** *Int J Gen Med.* 2010; **3**: 225–230.
PubMed Abstract | Publisher Full Text | Free Full Text

Ismail S, Farrands A, Wooding S, *et al.*: **Evaluating grant peer review in the health sciences a review of the literature.** Santa Monica, CA: RAND. 2009.
Reference Source

Jang D, Doh S, Kang GM, *et al.*: **Impact of Alumni Connections on Peer Review Ratings and Selection Success Rate in National Research.** *Sci Technol Hum Val.* 2016; **42**(1): 116–143.
Publisher Full Text

Jayasinghe UW, Marsh HW, Bond N: **Peer review in the funding of research in higher education: The Australian experience.** *Educational Evaluation and Policy Analysis.* 2001; **23**(4): 343–364.
Publisher Full Text

Jayasinghe UW, Marsh HW, Bond N: **A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings.** *J Roy Stat Soc.* 2003; **166**(3): 279–300.
Publisher Full Text

Kaatz A, Gutierrez B, Carnes M: **Threats to objectivity in peer review: the case of gender.** *Trends Pharmacol Sci.* 2014; **35**(8): 371–373.
PubMed Abstract | Publisher Full Text | Free Full Text

Kaatz A, Magua W, Zimmerman DR, *et al.*: **A quantitative linguistic analysis of National Institutes of Health R01 application critiques from investigators at one institution.** *Acad Med.* 2015; **90**(1): 69–75.
PubMed Abstract | Publisher Full Text | Free Full Text

Kaltman JR, Evans FJ, Danthi NS, *et al.*: **Prior publication productivity, grant percentile ranking, and topic-normalized citation impact of NHLBI cardiovascular R01 grants.** *Circ Res.* 2014; **115**(7): 617–624.
PubMed Abstract | Publisher Full Text | Free Full Text

Kaplan D: **How to improve peer review at N.I.H.** *Scientist.* 2005; 10.
Reference Source

Kotchen TA, Lindquist T, Malik K, *et al.*: **NIH peer review of grant applications for clinical research.** *JAMA.* 2004; **291**(7): 836–843.
PubMed Abstract | Publisher Full Text

Langfeldt L: **The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome.** *Soc Stud Sci.* 2001; **31**(6): 820–841.
Publisher Full Text

Langfeldt L: **The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments.** *Res Eval.* 2006; **15**(1): 31–41.
Publisher Full Text

Lauer MS, Danthi NS, Kaltman J, *et al.*: **Predicting Productivity Returns on Investment: Thirty Years of Peer Review, Grant Funding, and Publication of Highly Cited Papers at the National Heart, Lung, and Blood Institute.** *Circ Res.* 2015; **117**(3): 239–243.
PubMed Abstract | Publisher Full Text | Free Full Text

Lee CJ: **Commensuration Bias in Peer Review.** *Philos Sci.* 2015; **82**(5): 1272–1283.
Publisher Full Text

Lee CJ, Cassidy R, Sugimotot GZ, *et al.*: **Bias in peer review.** *Adv Inform Sci.* 2012; **64**(1): 2–17.
Publisher Full Text

Lee M, Om K, Koh J: **The bias of sighted reviewers in research proposal evaluation: a comparative analysis of blind and open review in Korea.** *Scientometrics.* 2000; **48**(1): 99–116.
Publisher Full Text

Li D: **Expertise vs. Bias in Evaluation: Evidence from the NIH.** *Harvard Business School Working Paper*, (No. 16-053). 2015.
Reference Source

Lindner MD, Vancea A, Chen MC, *et al.*: **NIH Peer Review: *Scored Review Criteria and Overall Impact*.** *Am J Eval.* 2016; **37**(2): 238–249.
PubMed Abstract | Publisher Full Text | Free Full Text

Linton JD: **Improving the Peer review process: Capturing more information and**

enabling high-risk/high-return research. *Res Policy.* 2016; **45**(9): 1936–1938.
Publisher Full Text

Lobb R, Petermann L, Manafo E, *et al.*: **Networking and knowledge exchange to promote the formation of transdisciplinary coalitions and levels of agreement among transdisciplinary peer reviewers.** *J Public Health Manag Pract.* 2013; **19**(1): E9–20.
PubMed Abstract | Publisher Full Text

Luukkonen T: **Conservatism and risk-taking in peer review: Emerging ERC practices.** *Res Eval.* 2012; **21**(1): 48–60.
Publisher Full Text

Mansilla V: **Assessing expert interdisciplinary work at the frontier: an empirical exploration.** *Res Eval.* 2006; **15**(1): 17–29.
Publisher Full Text

Marsh HW, Jayasinghe UW, Bond NW: **Gender differences in peer reviews of grant applications: A substantive-methodological synergy in support of the null hypothesis model.** *J Informetr.* 2011; **5**(1): 167–180.
Publisher Full Text

Martin MR, Kopstein A, Janice JM: **An analysis of preliminary and post-discussion priority scores for grant applications peer reviewed by the Center for Scientific Review at the NIH.** *PLoS One.* 2010; **5**(11): e13526.
PubMed Abstract | Publisher Full Text | Free Full Text

Mervis J: **Research grants. A radical change in peer review.** *Science.* 2014; **345**(6194): 248–249.
PubMed Abstract | Publisher Full Text

Mervis J: **U.S. RESEARCH MANAGEMENT. NSF tries two-step review, drawing praise--and darts.** *Science.* 2016; **353**(6299): 528–529.
PubMed Abstract | Publisher Full Text

Miner JT: **Behind Door #3: The Hard-to-Please Grant Reviewer.** *Res Manage Rev.* 2011; **18**(2): 85–107.
Reference Source

Morris ZS, Wooding S, Grant J: **The answer is 17 years, what is the question: understanding time lags in translational research.** *J R Soc Med.* 2011; **104**(12): 510–520.
PubMed Abstract | Publisher Full Text | Free Full Text

Mutz R, Bornmann L, Daniel HD: **Does Gender Matter in Grant Peer Review?: An Empirical Investigation Using the Example of the Austrian Science Fund.** *Z Psychol.* 2012; **220**(2): 121–129.
PubMed Abstract | Publisher Full Text | Free Full Text

NIH: **2007–2008 Peer Review Self-Study: final draft.** 2008.
Reference Source

Peckham D, Bosompra KN, Manuel C: **Evaluation of the Open Operating Grant Program, Final Report.** Ottawa: 2012.
Reference Source

Pier EL, Raclaw J, Nathan MJ, *et al.*: **Studying the study section: How group decision making in person and via videoconferencing affects the grant peer review process.** 2015.
Reference Source

Pier EL, Raclaw J, Nathan MJ, *et al.*: **Studying the study section: How group decision making in person and via videoconferencing affects the grant peer review process.** Working Paper No. 2015-6. Madison: Wisconsin Center for Education Research; 2015.
Reference Source

Pina DG, Hren D, Marusic A: **Peer Review Evaluation Process of Marie Curie Actions under EU's Seventh Framework Programme for Research.** *PLoS One.* 2015; **10**(6): e0130753.
PubMed Abstract | Publisher Full Text | Free Full Text

Porter AL, Rossini FA: **Peer Review of Interdisciplinary Research Proposals.** *Sci Technol Human Values.* 1985; **10**(3): 33–38.
Publisher Full Text

Reinhart M: **Peer review of grant applications in biology and medicine. Reliability, fairness, and validity.** *Scientometrics.* 2009; **81**(3): 789–809.
Publisher Full Text

Research Councils UK: **Efficiency and Effectiveness of Peer Review.** 2006.
Reference Source

RIN: **Peer review: A guide for researchers.** As of 5 January 2017, 2010.
Reference Source

Roorda S: **The real cost of the NSERC peer review is less than 5% of a proposed baseline grant.** *Account Res.* 2009; **16**(4): 229–231.
PubMed Abstract | Publisher Full Text

Roy R: **Funding science: the real defects of peer review and an alternative to it.** *Sci Technol Human Values.* 1985; **10**(3): 73–81.
Publisher Full Text

Sattler DN, McKnight PE, Naney L, *et al.*: **Grant Peer Review: Improving Inter-Rater Reliability with Training.** *PLoS One.* 2015; **10**(6): e0130450.
PubMed Abstract | Publisher Full Text | Free Full Text

Scarpa T: **Research funding: Peer Review at NIH.** *Science.* 2006; **311**(5757): 41.
PubMed Abstract | Publisher Full Text

Schroter S, Groves T, Højgaard L: **Surveys of current status in biomedical science grant review: funding organisations' and grant reviewers' perspectives.** *BMC Med.* 2010; **8**: 62.
PubMed Abstract | Publisher Full Text | Free Full Text

Snell RR: **Menage a quoi? Optimal number of peer reviewers.** *PLoS One.* 2015; **10**(4): e0120838.
PubMed Abstract | Publisher Full Text | Free Full Text

Spiegel AM: **Commentary: new guidelines for NIH peer review: improving the system or undermining it?** *Acad Med.* 2010; **85**(5): 746–748.
PubMed Abstract | Publisher Full Text

Spier RE: **Peer review and innovation.** *Sci Eng Ethics.* 2002; **8**(1): 99–108.
PubMed Abstract | Publisher Full Text

Street J, Baum F, Anderson IP: **Is peer review useful in assessing research proposals in Indigenous health? A case study.** *Health Res Policy Syst.* 2009; **7**: 2.
PubMed Abstract | Publisher Full Text | Free Full Text

Tamblyn R, McMahon M, Girard N, *et al.*: **Health services and policy research in the first decade at the Canadian Institutes of Health Research.** *CMAJ Open.* 2016; **4**(2): E213–21.
PubMed Abstract | Publisher Full Text | Free Full Text

Taylor M: **Of molecules, mice, and men: the relationship of biological complexity of research model to final rating in the grant peer review process of the Heart and Stroke Foundation of Canada.** *Paper presented at the Fourth International Congress on Peer Review in Biomedical Publication.* Barcelona, Spain. 2001.

Travis GDL, Collins HM: **New light on old boys: cognitive and institutional Particularism in the peer review system.** *Sci Technol Hum Val.* 1991; **16:** 322–341.
Publisher Full Text

Treasury HM: **Science and Innovation Investment Framework 2004 – 2014: Next Steps**. HMSO. 2006.
Reference Source

Turner S, Davidson P, Stanton L, *et al.*: **Features of successful bids for funding of applied health research: a cohort study.** *Health Res Policy Syst.* 2014; **12**: 54.
PubMed Abstract | Publisher Full Text | Free Full Text

Van Arensbergen P, van der Weijden I, van den Besselaar P: **Academic talent selection in grant review panels.** In: Prpic K, Van der Weijden I, Aseulova N. (Eds.) *(Re)searching Scientific Careers.* St. Petersburg: IHST/RAS & SSTNET/ESA; 2013.
Reference Source

van den Besselaar P: **Selection committee membership: Service or self-service.** *J Informetr.* 2012; **6**(4): 580–585.
Publisher Full Text

van den Besselaar P, Sandström U: **Early career grants, performance, and careers: A study on predictive validity of grant decisions.** *J Informetr.* 2015; **9**(4): 826–838.
Publisher Full Text

van der Lee R, Ellemers N: **Gender contributes to personal research funding success in The Netherlands.** *Proc Natl Acad Sci U S A.* 2015; **112**(40): 12349–12353.
PubMed Abstract | Publisher Full Text | Free Full Text

Viner N, Powell P, Green R: **Institutionalized biases in the award of research grants: a preliminary analysis revisiting the principle of accumulative advantage.** *Res Policy.* 2004; **33**(3): 443–454.
Publisher Full Text

Volker B, Steenbeek W: **No evidence that gender contributes to personal research funding success in The Netherlands: A reaction to van der Lee and Ellemers.** *Proc Natl Acad Sci U S A.* 2015; **112**(51): E7036–7037.
PubMed Abstract | Publisher Full Text | Free Full Text

Wang Q, Sandström U: **Defining the role of cognitive distance in the peer review process with an explorative study of a grant scheme in infection biology.** *Res Eval.* 2015; **24**(3): 271–281.
Publisher Full Text

Wenneras C, Wold A: **Nepotism and sexism in peer-review.** *Nature.* 1997; **387**(6631): 341–343.
PubMed Abstract | Publisher Full Text

Wooding S, Grant J: **Assessing Research: the Researchers' View.** Santa Monica, CA: 2003.
Reference Source

Open Peer Review

Current Referee Status: ? ✓

Version 1

Referee Report 14 August 2017

doi:10.5256/f1000research.12882.r24828

✓ **Aled Edwards**
Structural Genomics Consortium, University of Toronto, Toronto, ON, Canada

This is a nice and comprehensive review of the studies that assess the strengths and weaknesses of the peer review process.

Although there are no new conclusions, it does reveal or highlight the three major issues that plague our mechanism to dole out research funding and the use of peer review.

The first is fundamental - without a clear and agreed upon definition of what constitutes the "best" science, or the "best" outcomes, it will be impossible to know whether the grant adjudication process is meeting, or can ever meet, its objectives.

Second, the paper highlights what is often overlooked, and this is the real cost in time of writing funding proposals. Although scientists often manufacture a narrative that writing grants is good for them (for example, making you catch up on the literature), the real costs to the system are rarely factored in. I think this paper did a nice job in highlighted this issue.

Finally, I would have liked a greater discussion on what is an amazing disconnect. All the empirical evidence highlights the deficiencies of the process to allocate grant funding. It is clear that it is neither scientifically founded, nor evidence-based. Yet one of the only strongly supported aspects of the peer review process is that is has the strong support of the community. I find this fascinating.

Is the topic of the review discussed comprehensively in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Is the review written in accessible language?
Yes

Are the conclusions drawn appropriate in the context of the current research literature?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 09 August 2017

doi:10.5256/f1000research.12882.r24827

Adrian Barnett
Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Qld, Australia

This is a very useful review of an area that is vitally important for science and that is constantly being examined by funding agencies. It included some papers that I had not read, but there were a few additional papers that I thought could be included (detailed below).

The results once again highlight the incredible lack of studies in this area. The paper ends with some sensible recommendations, including the need for funders to experiment more and make their data available to researchers.

I was surprised that some of the more innovative solutions to funding peer review were not included, specifically using prediction markets [1] and using the "wisdom of the crowd"[2].

Why was 2009 chosen as the time threshold? Is it because that was the year of the previous review?

This paper should be included in the discussion on interdisciplinary research: Interdisciplinary research has consistently lower funding success, Lindell Bromham, Russell Dinnage & Xia Hua. Nature 534, 684–687 doi:10.1038/nature18315 [3]. This paper agrees with the other two mentioned, as there are lower success rates for applications with more cross-disciplinary researchers.

One force against "cognitive particularism" is that strict conflict of interest rules from funding bodies can often rule out reviewers with the greatest knowledge, particularly in small fields or small countries. This study may touch on this issue: Abdoul H *et al*, Non-Financial Conflicts of Interest in Academic Grant Evaluation: A Qualitative Study of Multiple Stakeholders in France, PLoS ONE, 7/4: e35247.[4]

In terms of using technology in the review process, some researchers have suggested that videos may produce more reliable peer reviewer ratings and take less time to prepare: Doran MR, Lott WB, Doran SE. Trends Biochem Sci. 2014 Apr;39(4):151-3. doi: 10.1016/j.tibs.2014.01.004. Multimedia: a necessary step in the evolution of research funding applications[5].

Minor comments

- Introduction, 1st paragraph. As an Australian researcher I would argue that the ERA has not really measured research quality, rather it has simply measured research output. Maybe you could say "Funders have attempted to gather evidence…"

- The ">95%" figure in the abstract feels about right, but is there a reference for this?

- For the Google search terms, “fellowship” could also have been added, so “Fellowship OR funding”.
- The link to this paper did not work: The Novelty Paradox & Bias for Normal Science: Evidence from Randomized Medical Grant Proposal Evaluations
- Page 6, “when reviewing research closer to application”, I didn’t understand this.
- Page 7, “only affects the funding decision for around 10 per cent of applications relative to original scores” but that could still be an important percentage, particularly if it’s those near the funding line
- Page 9, “found that panel assessments of full proposals and shorter anonymised versions of the same proposals showed weak correlations” do you need to add, “implying that knowledge of the applicants influences the score”. Although as well as a change in blinding there was also a change in the size of the application, so it may be hard to conclude anything about cronyism here.
- Footnote 9, the NZ Health Research Council has been using random allocation for this scheme since at least April 2015
- I agree that giving more feedback would improve the value of the process (page 12), but our experience with the NHMRC is that this also opens them up to appeals which can take a lot of time for their staff. This doesn’t mean that we shouldn’t try, and giving frank feedback was a feature of a funding scheme we designed (already cited paper “Streamlined research funding using short proposals and accelerated peer review: an observational study”).
- This paper may be of interest: Scientometrics, July 2016, Volume 108, Issue 1, pp 263–288, The consequences of competition: simulating the effects of research grant allocation strategies.[6]
- This parody of a grant application may be useful for the section on peer reviewers being biased against innovative proposals: doi:10.1097/ede.0000000000000453 “John Snow’s Grant Application”[7]
- This paper examined the costs of applying for NIH funding and could be included in the section on the costs to applicants: Nursing Outlook, Volume 63, Issue 6, November–December 2015, Pages 639-649, Time and costs of preparing and submitting an NIH grant application at a school of nursing, https://doi.org/10.1016/j.outlook.2015.09.003[8]

References

1. Munafo MR, Pfeiffer T, Altmejd A, Heikensten E, Almenberg J, Bird A, Chen Y, Wilson B, Johannesson M, Dreber A: Using prediction markets to forecast research evaluations.*R Soc Open Sci*. 2015; **2** (10): 150287 PubMed Abstract | Publisher Full Text
2. Bollen J, Crandall D, Junk D, Ding Y, Börner K: An efficient system to fund science: from proposal review to peer-to-peer distributions. *Scientometrics*. 2017; **110** (1): 521-528 Publisher Full Text
3. Bromham L, Dinnage R, Hua X: Interdisciplinary research has consistently lower funding success. *Nature*. 2016; **534** (7609): 684-687 Publisher Full Text
4. Abdoul H, Perrey C, Tubach F, Amiel P, Durand-Zaleski I, Alberti C: Non-financial conflicts of interest in academic grant evaluation: a qualitative study of multiple stakeholders in France.*PLoS One*. 2012; **7** (4): e35247 PubMed Abstract | Publisher Full Text

5. Doran MR, Lott WB, Doran SE: Multimedia: a necessary step in the evolution of research funding applications.*Trends Biochem Sci*. 2014; **39** (4): 151-3 PubMed Abstract | Publisher Full Text
6. Höylä T, Bartneck C, Tiihonen T: The consequences of competition: simulating the effects of research grant allocation strategies. *Scientometrics*. 2016; **108** (1): 263-288 Publisher Full Text
7. Rothman K: John Snow's Grant Application. *Epidemiology*. 2016; **27** (3): 311-313 Publisher Full Text
8. Kulage KM, Schnall R, Hickey KT, Travers J, Zezulinski K, Torres F, Burgess J, Larson EL: Time and costs of preparing and submitting an NIH grant application at a school of nursing.*Nurs Outlook*. **63** (6): 639-49 PubMed Abstract | Publisher Full Text

Is the topic of the review discussed comprehensively in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Partly

Is the review written in accessible language?
Yes

Are the conclusions drawn appropriate in the context of the current research literature?
Yes

Competing Interests: In the past few years I have had one Skype conversation and around 10 email conversations with Steve Wooding about the peer review process for funding applications.

Referee Expertise: Statistics

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.